12-OST-09980729.032PCT

JC03 Rec'd PCT/PTO 19 OCT 2001

IPE
OCT 19 2001
TRADEMARK OFFICE

FORM PTO-1390 U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE	ATTORNEY'S DOCKET NUMBER
TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371	**PF-0687 USN**
	U.S. APPLICATION NO. (If known, see 37 CFR 1.5) TO BE ASSIGNED **09/980729**

INTERNATIONAL APPLICATION NO.	INTERNATIONAL FILING DATE	PRIORITY DATE CLAIMED
PCT/US00/10882	**20 April 2000**	**21 April 1999**

TITLE OF INVENTION
CARBOHYDRATE-MODIFYING ENZYMES

APPLICANT(S) FOR DO/EO/US
INCYTE Genomics, INC.; LAL, Preeti; YUE, Henry; TANG, Y. Tom; HILLMAN, Jennifer L.; BAUGHN, YAGN, Junming

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. ☒ This is the **FIRST** submission of items concerning a filing under 35 U.S.C. 371.
2. ☐ This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.
3. ☐ This is an express request to promptly begin national examination procedures (35 U.S.C. 371 (f)).
4. ☐ The US has been elected by the expiration of 19 months from the priority date (PCT Article 31).
5. ☒ A copy of the International Application as filed (35 U.S.C. 371(c)(2))
   a. ☐ is attached hereto (required only if not communicated by the International Bureau)
   b. ☐ has been communicated by the International Bureau.
   c. ☒ is not required, as the application was filed in the United States Receiving Office (RO/US).
6. ☐ An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
7. ☒ Amendments to the claims of the International Application under PCT Article 19 (35 U.S.C. 371(c)(3))
   a. ☐ are attached hereto (required only if not communicated by the International Bureau).
   b. ☐ have been communicated by the International Bureau.
   c. ☐ have not been made; however, the time limit for making such amendments has NOT expired.
   d. ☒ have not been made and will not be made.
8. ☐ An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371(c)(3)).
9. ☒ An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).
10. ☐ An English language translation of the annexes to the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 16 below concern document(s) or information included:

11. ☐ An Information Disclosure Statement under 37 CFR 1.97 and 1.98.
12. ☒ **An assignment document for recording. A separate cover sheet in compliance with** 37 CFR 3.27 and 3.31 is included.
13. ☒ A FIRST preliminary amendment, as follows:

Cancel in this application original claims #18, 21, and 22 before calculating the filing fee, without prejudice or disclaimer. Applicants submit that these claims were included in the application as filed in the interest of providing notice to the public of certain specific subject matter intended to be claimed, and are being canceled at this time in the interest of reducing filing costs. Applicants expressly state that these claims are not being canceled for reasons related to patentability, and are in fact fully supported by the specification as filed. Applicants expressly reserve the right to reinstate these claims or to add other claims during prosecution of this application or a continuation or divisional application. Applicants expressly do not disclaim the subject matter of any invention disclosed herein which is not set forth in the instantly filed claims.

☐ A SECOND or SUBSEQUENT preliminary amendment.

14. ☐ A substitute specification.
15. ☐ A change of power of attorney and/or address letter.
16. ☒ Other items or information:

1) Transmittal Letter (2 pp, in duplicate)
2) Return Postcard
3) Express Mail Label No.: **EL 856 148741 US**
4) Request to Transfer (2 pp, in duplicate)

09980729 .032802
JC13 Rec'd PCT/PTO 19 OCT 2001

U.S. APPLICATION NO. (if known, see 37 CFR 1.5) TO BE ASSIGNED 09/980729	INTERNATIONAL APPLICATION NO.: PCT/US00/10882	ATTORNEY'S DOCKET NUMBER PF-0687 USN

17. ☒ The following fees are submitted:

BASIC NATIONAL FEE (37 CFR 1.492(a)(1)-(5):

Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO.....$1000.00

☐International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO..$860.00

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO.....$710.00

☒**International preliminary examination fee paid to USPTO (37 CFR 1.482) but all claims did not satisfy provisions of PCT Article 33(1)-(4).....$740.00**

☐International preliminary examination fee paid to USPTO (37 CFR 1.482) and all claims satisfied provisions of PCT Article 33(1)-(4).....$100.00

ENTER APPROPRIATE BASIC FEE AMOUNT =				$740.00	
Surcharge of **$130.00** for furnishing the oath or declaration later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1.492(e)).				$	
CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE		
Total Claims	20 =	0	X **$ 18.00**	$	
Independent Claims	2 =	0	X **$ 80.00**	$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			**+ $270.00**	$	
TOTAL OF ABOVE CALCULATIONS =				$	
☐ Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2.				$	
= **SUBTOTAL**				$740.00	
Processing fee of **$130.00** for furnishing the English translation later than ☐ 20 ☐ 30 months from the earliest claimed priority date (37 CFR 1492(f)). +				$	
TOTAL NATIONAL FEE =				$740.00	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by the appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +				$40.00	
TOTAL FEES ENCLOSED =				$780.00	
				Amount to be Refunded:	$
				Charged:	$

a. ☐ A check in the amount of $_____ to cover the above fees is enclosed.

b. ☒ Please charge my Deposit Account No. 09-0108 in the amount of $780.00 to cover the above fees.

c. ☒ The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 09-0108. A duplicate copy of this sheet is enclosed.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137(a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

INCYTE GENOMICS, INC.
3160 Porter Drive
Palo Alto, CA 94304

SIGNATURE

NAME: Richard C. Ekstrom

REGISTRATION NUMBER: 37,027

DATE: 19 October 2001

DT16 Rec'd PCT/PTO JUN 14 2002

Docket No.: PF-0687 USN

Certificate of Mailing

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: **Box PCT**, Commissioner for Patents, Washington, D.C. 20231 on June 6, 2002.

By: Joyce Abriam Printed: Joyce Abriam

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Lal et al.

Title: CARBOHYDRATE-MODIFYING ENZYME

Serial No.: 09/980,729 Filing Date: To Be Assigned

Examiner: To Be Assigned Group Art Unit: To Be Assigned

Box PCT
Commissioner for Patents
Washington, D.C. 20231

TRANSMITTAL FEE SHEET

Sir:

Transmitted herewith are the following for the above-identified application:

1. Return Receipt Postcard;
2. Response to Notification of Missing Requirements under 35 USC § 371 (2 pp.);
3. Copy of Notification Missing Requirements under 35 USC § 371 dated 4/26/2002 (3 pp.);
4. Courtesy copy of documents filed on February 15, 2002 (16 pp);
5. Substitute Submission Under 37 CFR § 1.821-1.825 Sequence Listing (1 pg.); and
6. One (1) substitute diskette containing the Sequence Listing.

The fee has been calculated as shown below.

X No additional Fee is required.

___ Please charge Deposit Account No. **09-0108** in the amount of : $ _____

Applicants believe that no fee is due with this paper. However, if the Commissioner is hereby authorized to charge any additional fees required under 37 CFR 1.16 and 1.17, or credit overpayment to Deposit Account No. 09-0108. A duplicate copy of this sheet is enclosed.

Respectfully submitted,

INCYTE GENOMICS, INC.

Date: 6 June 2002

Diana Hamlet-Cox
Reg. No. 33,302
Direct Dial Telephone: (650) 845-4639

3160 Porter Drive
Palo Alto, California 94304
Phone: (650) 855-0555
Fax: (650) 845-4166

CARBOHYDRATE-MODIFYING ENZYMES

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of carbohydrate-modifying enzymes and to the use of these sequences in the diagnosis, treatment, and prevention of carbohydrate metabolism disorders, autoimmune/inflammatory disorders, and cancer.

BACKGROUND OF THE INVENTION

Carbohydrates, including sugars or saccharides, starch, and cellulose, are aldehyde or ketone compounds with multiple hydroxyl groups. The importance of carbohydrate metabolism is demonstrated by the sensitive regulatory system in place for maintenance of blood glucose levels. Two pancreatic hormones, insulin and glucagon, promote increased glucose uptake and storage by cells and increased glucose release from cells, respectively. Carbohydrates have three important roles in mammalian cells. First, carbohydrates are used as energy stores, fuels, and metabolic intermediates. Carbohydrates are broken down to form energy in glycolysis and are stored as glycogen for later use. Second, the sugars deoxyribose and ribose form part of the structural support of DNA and RNA, respectively. Third, carbohydrate modifications are added to secreted and membrane proteins and lipids as they traverse the secretory pathway. Indeed, 2-10% of the content of eukaryotic cell membranes are contributed by oligosaccharides on membrane glycoproteins and glycolipids. Oligosaccharide modifications of carbohydrates create great structural diversity. Modifications on glycoproteins and glycolipids are mostly located on the extracellular side of the plasma membrane and are important for intercellular recognition (Stryer, L. (1988) Biochemistry, W.H. Freeman and Company, New York NY, pp. 298-299, 331-347).

N- and O-linked oligosaccharides are transferred to proteins and modified in a series of enzymatic reactions that occurs in the endoplasmic reticulum (ER) and Golgi. Oligosaccharides stabilize a protein during and after folding, orient the protein in the membrane, improve the protein's solubility, and act as a signal for lysosome targeting. Glycolipids, along with phospholipids and cholesterol, form the membrane of cells. Examples of glycolipids include blood group antigens on erythrocytes and gangliosides in the myelin sheath of neurons (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York NY, pp. 612-615).

Carbohydrates also form glycosaminoglycans (GAGs), which are linear unbranched polysaccharides composed of repetitive disaccharide units. GAGs exist free or as part of proteoglycans, large molecules composed of a core protein attached to one or more GAGs. GAGs are found on the cell surface, inside cells, and in the extracellular matrix. The GAG hyaluronan is abundant in synovial fluid (Pitsillides, A.A. et al. (1993) Int. J. Exp. Pathol. 74:27-34).

Proteoglycans in the extracellular matrix of connective tissues such as cartilage are essential for distributing the load in weight-bearing joints. Cell-surface-attached proteoglycans anchor cells to the extracellular matrix. Both extracellular and cell-surface proteoglycans bind growth factors, facilitating their binding to cell-surface receptors and subsequent triggering of signal transduction pathways (Lodish, supra, pp. 1139-1142).

Man_9-mannosidase is an α 1,2-mannosidase (glycosyl hydrolase) involved in the early processing of N-linked oligosaccharides. This enzyme catalyzes the specific cleavage of α 1,2-mannosidic linkages in Man_9-$(GlcNAc)_2$ and Man_5-$(GlcNAc)_2$. Multiple α 1,2-mannosidases have been identified in mammalian cells and may be needed for the processing of distinct classes of N-glycoproteins. Man_9-mannosidase is a Type II membrane protein with a short cytoplasmic tail, a single transmembrane domain, and a large luminal catalytic domain. The pig liver Man_9-mannosidase is localized to the ER and transient vesicles while the human kidney Man_9-mannosidase is localized to the Golgi (Bause, E. et al. (1993) Eur. J. Biochem. 217:535-540; Bieberich, E. and E. Bause (1995) Eur. J. Biochem. 233:644-649).

Transferases participate in reactions essential to the synthesis and degradation of cellular components such as carbohydrates. For example, galactosyltransferase catalyzes the reaction producing galactose beta-1,4-N-acetylglucosamine, has a role in lactose synthesis, and, as a component of the plasma membrane, may function in intracellular recognition and/or adhesion (Masri, K.A. et al. (1988) Biochem. Biophys. Res. Commun. 157:657-663).

Synthetases are another class of carbohydrate-modifying enzymes that have critical roles in proper cell functioning. For example, synthesis of sialylated glycoconjugates requires the synthesis of cytidine 5'-monophosphate N-acetylneuraminic acid (CMP-Neu5Ac), a reaction catalyzed by CMP-Neu5Ac synthetase (Munster, A.K. et al. (1998) Proc. Natl. Acad. Sci. USA 95:9140-9145). Sialic acids of cell surface glycoproteins and glycolipids contribute to proper structure and function in a variety of tissues.

Another class of carbohydrate-modifying enzymes is the glucosidases that catalyze the release of glucose from carbohydrates through hydrolysis of the glycosidic link in various glucosides. The inherited disorder type I Gaucher disease, characterized by hematologic abnormalities, can be detected in a heterozygous or homozygous individual through an assay of leukocyte beta-glucosidase levels (Raghavan, S.S. et al. (1980) Am. J. Hum. Genet. 32:158-173).

Carbohydrate metabolism is altered in several other disorders. Diabetes mellitus is characterized by abnormally high blood glucose (hyperglycemia). Type I diabetes results from an autoimmune-related loss of pancreatic insulin-secreting cells. Type II diabetes results from insulin resistance and impaired insulin secretory response to glucose, and is associated with obesity. Hypoglycemia, or abnormally low blood glucose levels, has several causes including drug use,

genetic deficiencies in carbohydrate metabolism enzymes, cancer, liver disease, and renal disease (Berkow, R. et al. (1992) The Merck Manual of Diagnosis and Therapy, Internet Edition, Section 8, Chapter 91, Diabetes Mellitus, Hypoglycemia).

Changes in glycosaminoglycan (GAG) levels are associated with several autoimmune diseases. Both increases and decreases in various GAGs occur in patients with autoimmune thyroid disease and autoimmune diabetes mellitus. Antibodies to GAGs were found in patients with systemic lupus erythematosus and autoimmune thyroid disease (Hansen, C. et al. (1996) Clin. Exp. Rheum. 14 (Suppl. 15):S59-S67).

Carbohydrate metabolism is associated with cancer. Reduced GAG and proteoglycan expression is associated with human lung carcinomas (Nackaerts, K. et al. (1997) Int. J. Cancer 74:335-345). The carbohydrate determinants, sialyl Lewis A and sialyl Lewis X, are frequently expressed on human cancer cells. These determinants, ligands for the cell adhesion molecule E-selectin, are involved in the adhesion of cancer cells to vascular endothelium and contribute to hematogenous metastasis of cancer (Kannagi, R. (1997) Glycoconj. J. 14:577-584). Alterations of the N-linked carbohydrate core structure of cell surface glycoproteins are linked to colon and pancreatic cancers (Schwarz, R.E. et al. (1996) Cancer Lett. 107:285-291). Reduced expression of the Sda blood group carbohydrate structure in cell surface glycolipids and glycoproteins is observed in gastrointestinal cancer (Dohi, T. et al. (1996) Int. J. Cancer 67:626-631).

The discovery of new carbohydrate-modifying enzymes and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of carbohydrate metabolism disorders, autoimmune/inflammatory disorders, and cancer.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, carbohydrate-modifying enzymes, referred to collectively as "CME" and individually as "CME-1," "CME-2," "CME-3," "CME-4," and "CME-5." In one aspect, the invention provides an isolated polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-5.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally

occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. In one alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:6-10.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

The invention further provides an isolated polynucleotide comprising a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, c) a polynucleotide sequence complementary to a), or d) a polynucleotide sequence complementary to b). In one alternative, the polynucleotide comprises

at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10, c) a polynucleotide sequence complementary to a), or d) a polynucleotide sequence complementary to b). The method comprises a) hybridizing the sample with a probe comprising at least 16 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 30 contiguous nucleotides. In another alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a pharmaceutical composition comprising an effective amount of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and a pharmaceutically acceptable excipient. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional CME, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional CME, comprising administering to a patient in need of such treatment the pharmaceutical composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, or d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional CME, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:6-10, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length sequences encoding CME.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous sequences, and methods, algorithms, and searchable databases used for analysis of CME.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones encoding CME were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze CME, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"CME" refers to the amino acid sequences of substantially purified CME obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of CME. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CME either by directly interacting with CME or by acting on components of the biological pathway in which CME participates.

An "allelic variant" is an alternative form of the gene encoding CME. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding CME include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as CME or a

polypeptide with at least one functional characteristic of CME. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding CME, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding CME. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent CME. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of CME is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of CME. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of CME either by directly interacting with CME or by acting on components of the biological pathway in which CME participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind CME polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin,

and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic CME, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" and "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence "5' A-G-T 3'" bonds to the complementary sequence "3' T-C-A 5'." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acid strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding CME or fragments of CME may be

employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using the XL-PCR kit (Perkin-Elmer, Norwalk CT) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of one or more Incyte Clones and, in some cases, one or more public domain ESTs, using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that, when made, least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to the chemical modification of a polypeptide sequence, or a polynucleotide sequence. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "fragment" is a unique portion of CME or the polynucleotide encoding CME which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:6-10 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:6-10, for example, as distinct from any other sequence in the same genome. A fragment of SEQ ID NO:6-10 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:6-10 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:6-10 and the region of SEQ ID NO:6-10 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-5 is encoded by a fragment of SEQ ID NO:6-10. A fragment of SEQ ID NO:1-5 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-5. For example, a fragment of SEQ ID NO:1-5 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-5. The precise length of a fragment of SEQ ID NO:1-5 and the region of SEQ ID NO:1-5 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

The term "similarity" refers to a degree of complementarity. There may be partial similarity or complete similarity. The word "identity" may substitute for the word "similarity." A partially

complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially similar sequence or hybridization probe will compete for and inhibit the binding of a completely similar (identical) sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% similarity or identity). In the absence of non-specific binding, the substantially similar sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequence pairs.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2

Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the hydrophobicity and acidity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default

residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance.

The term "humanized antibody" refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of identity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about

1% (w/v) SDS, and about 100 µg/ml denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Generally, such wash temperatures are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of CME which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment

of CME which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of distinct polynucleotides on a substrate.

The terms "element" and "array element" in a microarray context, refer to hybridizable polynucleotides arranged on the surface of a substrate.

The term "modulate" refers to a change in the activity of CME. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of CME.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Probe" refers to nucleic acid sequences encoding CME, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the

specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel et al.,1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have

been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding CME, or fragments thereof, or CME itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

"Transformation" describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign

nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, and plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a

propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

THE INVENTION

The invention is based on the discovery of new human carbohydrate-modifying enzymes (CME), the polynucleotides encoding CME, and the use of these compositions for the diagnosis, treatment, or prevention of carbohydrate metabolism disorders, autoimmune/inflammatory disorders, and cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding CME. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which nucleic acids encoding each CME were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. The Incyte clones in column 5 were used to assemble the consensus nucleotide sequence of each CME and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding CME. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:6-10 and to distinguish between SEQ ID NO:6-10 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express CME as a fraction of total tissues expressing CME. Column 4 lists

diseases, disorders, or conditions associated with those tissues expressing CME as a fraction of total tissues expressing CME. Column 5 lists the vectors used to subclone each cDNA library.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding CME were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

The invention also encompasses CME variants. A preferred CME variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the CME amino acid sequence, and which contains at least one functional or structural characteristic of CME.

The invention also encompasses polynucleotides which encode CME. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:6-10, which encodes CME. The polynucleotide sequences of SEQ ID NO:6-10, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding CME. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding CME. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:6-10 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:6-10. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of CME.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding CME, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring CME, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode CME and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring CME under appropriately selected

conditions of stringency, it may be advantageous to produce nucleotide sequences encoding CME or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding CME and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode CME and CME derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding CME or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:6-10 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Perkin-Elmer), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Perkin-Elmer). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Perkin-Elmer), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding CME may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences,

such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin-Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode CME may be cloned in recombinant DNA molecules that direct expression of CME, or

fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express CME.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter CME-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of CME, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding CME may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, CME itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin-Elmer). Additionally, the amino acid sequence of CME, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY.)

In order to express a biologically active CME, the nucleotide sequences encoding CME or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding CME. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding CME. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding CME and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding CME and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding CME. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending

upon the use intended for polynucleotide sequences encoding CME. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding CME can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding CME into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of CME are needed, e.g. for the production of antibodies, vectors which direct high level expression of CME may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of CME. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of CME. Transcription of sequences encoding CME may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding CME may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses CME in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of CME in cell lines is preferred. For example, sequences encoding CME can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*⁻ and *apr*⁻ cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), ß glucuronidase and its substrate ß-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding CME is inserted within a marker gene sequence, transformed cells containing sequences encoding CME can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding CME under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates

expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding CME and that express CME may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of CME using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CME is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding CME include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding CME, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding CME may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode CME may be designed to contain signal sequences which direct secretion of CME through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the

inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding CME may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric CME protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of CME activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the CME encoding sequence and the heterologous protein sequence, so that CME may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled CME may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

Fragments of CME may be produced not only by recombinant means, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, supra, pp. 55-60.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the ABI 431A peptide synthesizer (Perkin-Elmer). Various fragments of

CME may be synthesized separately and then combined to produce the full length molecule.

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of CME and carbohydrate-modifying enzymes. In addition, the expression of CME is closely associated with cardiovascular, developmental, gastrointestinal, hematopoietic/immune, nervous, reproductive, and urologic tissues. Therefore, CME appears to play a role in carbohydrate metabolism disorders, autoimmune/inflammatory disorders, and cancers. In the treatment of disorders associated with increased CME expression or activity, it is desirable to decrease the expression or activity of CME. In the treatment of disorders associated with decreased CME expression or activity, it is desirable to increase the expression or activity of CME.

Therefore, in one embodiment, CME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CME. Examples of such disorders include, but are not limited to, a carbohydrate metabolism disorder such as diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, hypoglycemia, glucagonoma, galactosemia, hereditary fructose intolerance, fructose-1,6-diphosphatase deficiency, obesity, congenital type II dyserythropoietic anemia, mannosidosis, neuraminidase deficiency, galactose epimerase deficiency, glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, and inherited abnormalities of pyruvate metabolism; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing CME or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CME including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified CME in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CME including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of CME may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of CME including, but not limited to, those listed above.

In a further embodiment, an antagonist of CME may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CME. Examples of such disorders include, but are not limited to, those carbohydrate metabolism disorders, autoimmune/inflammatory disorders, and cancers described above. In one aspect, an antibody which specifically binds CME may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express CME.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding CME may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of CME including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of CME may be produced using methods which are generally known in the art. In particular, purified CME may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind CME. Antibodies to CME may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with CME or with any fragment or oligopeptide thereof

which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to CME have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of CME amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to CME may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce CME-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for CME may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and

easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between CME and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering CME epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for CME. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of CME-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple CME epitopes, represents the average affinity, or avidity, of the antibodies for CME. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular CME epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the CME-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of CME, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington, DC; Liddell, J.E. and Cryer, A. (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of CME-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding CME, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding CME may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to

polynucleotides encoding CME. Thus, complementary molecules or fragments may be used to modulate CME activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding CME.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors to express nucleic acid sequences complementary to the polynucleotides encoding CME. (See, e.g., Sambrook, supra; Ausubel, 1995, supra.)

Genes encoding CME can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding CME. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding CME. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may be employed. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding CME.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides,

corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding CME. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of CME, antibodies to CME, and mimetics, agonists, antagonists, or inhibitors of CME. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including,

but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of CME, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example CME or fragments thereof, antibodies of CME, and agonists, antagonists or inhibitors of CME, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by

standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind CME may be used for the diagnosis of disorders characterized by expression of CME, or in assays to monitor patients being treated with CME or agonists, antagonists, or inhibitors of CME. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for CME include methods which utilize the antibody and a label to detect CME in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring CME, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of CME expression. Normal or standard values for CME expression are established by combining body fluids or cell extracts

taken from normal mammalian subjects, for example, human subjects, with antibody to CME under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of CME expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding CME may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of CME may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of CME, and to monitor regulation of CME levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding CME or closely related molecules may be used to identify nucleic acid sequences which encode CME. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding CME, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the CME encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:6-10 or from genomic sequences including promoters, enhancers, and introns of the CME gene.

Means for producing specific hybridization probes for DNAs encoding CME include the cloning of polynucleotide sequences encoding CME or CME derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding CME may be used for the diagnosis of disorders associated with expression of CME. Examples of such disorders include, but are not limited to, a carbohydrate metabolism disorder such as diabetes, insulin-dependent diabetes mellitus, non-insulin-dependent diabetes mellitus, hypoglycemia, glucagonoma, galactosemia, hereditary fructose intolerance, fructose-1,6-diphosphatase deficiency, obesity, congenital type II dyserythropoietic anemia, mannosidosis, neuraminidase deficiency, galactose epimerase deficiency,

glycogen storage diseases, lysosomal storage diseases, fructosuria, pentosuria, and inherited abnormalities of pyruvate metabolism; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding CME may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered CME expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding CME may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding CME may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding CME in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of CME, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a

fragment thereof, encoding CME, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding CME may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding CME, or a fragment of a polynucleotide complementary to the polynucleotide encoding CME, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

Methods which may also be used to quantify the expression of CME include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene

function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding CME may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding CME on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, CME, its catalytic or immunogenic fragments, or

oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between CME and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with CME, or fragments thereof, and washed. Bound CME is then detected by methods well known in the art. Purified CME can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding CME specifically compete with a test compound for binding CME. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with CME.

In additional embodiments, the nucleotide sequences which encode CME may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in particular U.S. Ser. No. 60/130,383, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA

purity. In some cases, RNA was treated with DNase. For most libraries, poly(A+) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Pharmaceuticals, Palo Alto CA). Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Life Technologies.

II. Isolation of cDNA Clones

Plasmids were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

cDNA sequencing reactions were processed using standard methods or high-throughput

instrumentation such as the ABI CATALYST 800 (Perkin-Elmer) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Perkin-Elmer) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programing, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length

amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:6-10. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST were used to search for identical or related molecules in nucleotide databases such as GenBank or LIFESEQ (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding CME occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

V. Extension of CME Encoding Polynucleotides

The full length nucleic acid sequences of SEQ ID NO:6-10 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and β-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham

Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Perkin-Elmer).

In like manner, the nucleotide sequences of SEQ ID NO:6-10 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for such extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:6-10 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the CME-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring CME. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of CME. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the CME-encoding transcript.

IX. Expression of CME

Expression and purification of CME is achieved using bacterial or virus-based expression systems. For expression of CME in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express CME upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of CME in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding CME by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, CME is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from CME at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified CME obtained by these methods can be used directly in the following activity assay.

X. Demonstration of CME Activity

Galactosyltransferase activity in CME is determined by measuring the transfer of galactose from UDP-galactose to a GlcNAc-terminated oligosaccharide chain in a radioactive assay (Hennet, T. et al. (1998) J. Biol. Chem. 273:58-65). An aliquot of CME is incubated with 14 µl of assay stock solution (180 mM sodium cacodylate, pH 6.5, 1 mg/ml bovine serum albumin, 0.26 mM UDP-galactose, 2 µl of UDP-[^{3}H]galactose), 1 µl of $MnCl_2$ (500 mM), and 2.5 µl of GlcNAcβO-$(CH_2)_8$-CO_2Me (37 mg/ml in dimethyl sulfoxide) for 60 minutes at 37°C. The reaction is quenched by the addition of 1 ml of water and loaded on a C18 Sep-Pak cartridge (Waters), and the column is washed twice with 5 ml of water to remove unreacted UDP-[^{3}H]galactose. The [^{3}H]galactosylated GlcNAcβO-$(CH_2)_8$-CO_2Me remains bound to the column during the water washes and is eluted with 5 ml of methanol. Radioactivity in the eluted material is measured by liquid scintillation counting and is proportional to CME galactosyltransferase activity.

Mannosidase activity in CME is demonstrated by the ability to release mannose from $Man_9(GlcNAc)_2$ oligosaccharide (Schweden, J. et al. (1986) Eur. J. Biochem. 157:563-570). CME, in 200 mM phosphate buffer, pH 6.5 and 1% Triton X-100, is mixed with $[^{14}C](Man_9)(GlcNAc)_2$ (2-3 x 10^3 cpm) in a final volume of 30 µl at 37°C for 60 minutes. The reaction is terminated by the addition of 30 µl glacial acetic acid. The amount of liberated [^{14}C]mannose, analyzed by paper chromatography in 2-propanol/acetic acid/water (29/4/9, by volume), is proportional to the activity of CME in the starting sample.

XI. Functional Assays

CME function is assessed by expressing the sequences encoding CME at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT plasmid (Life Technologies) and pCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 µg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of CME on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding CME and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art.

Expression of mRNA encoding CME and other genes of interest can be analyzed by northern analysis or microarray techniques.

XII. Production of CME Specific Antibodies

CME substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the CME amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Perkin-Elmer) using fmoc-chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-CME activity by, for example, binding the peptide or CME to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XIII. Purification of Naturally Occurring CME Using Specific Antibodies

Naturally occurring or recombinant CME is substantially purified by immunoaffinity chromatography using antibodies specific for CME. An immunoaffinity column is constructed by covalently coupling anti-CME antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing CME are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of CME (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/CME binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and CME is collected.

XIV. Identification of Molecules Which Interact with CME

CME, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled CME, washed, and any wells with labeled CME complex are assayed. Data obtained using different concentrations of CME are used to calculate values for the number, affinity, and association of CME with the candidate

molecules.

Alternatively, molecules interacting with CME are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	6	000422	U937NOT01	000422H1 (U937NOT01), 458113F1 (KERANOT01), 1271685F1 (TESTTUT02), 1725667T6 (PROSNOT14), 1752778H1 (LIVRTUT01), 2459639H1 (THYRNOT08), 2622986H1 (KERANOT02)
2	7	983984	TONGTUT01	715334R6 (PROSTUT01), 983984H1 (TONGTUT01), 983984R1 (TONGTUT01), 983984X301D1 (TONGTUT01), 3168253F6 (BRSTNOT18)
3	8	2210054	SINTFET03	077799R1 (SYNORAB01), 856950R1 (NGANNOT01), 1558537F1 (SPLNNOT04), 1654114F6 (PROSTUT08), 2210054H1 (SINTFET03), 3387285H1 (LUNGTUT17), 4559124H1 (KERATXT01)
4	9	2618358	GBLANOT01	1634052F6 (COLNNOT19), 1634052T6 (COLNNOT19), 2618358H1 (GBLANOT01), 2618358X300D1 (GBLANOT01), 2618358X303D1 (GBLANOT01), 2618358X313D1 (GBLANOT01), 2673370F6 (KIDNNOT19), 4072754H1 (KIDNNOT26)
5	10	2912330	KIDNTUT15	089378H1 (LIVRNOT01), 2531411H1 (GBLANOT02), 2912330H1 (KIDNTUT15), SZAF00178F1, SZAF00234F1, SZAF00054F1, SZAF00080F1

09960729 .032402

Table 2

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
1	434	S113 T92 T121 S317 S346 S376 S325 S336 S348 Y212 Y229	N214 N429	Signal peptide: M1-G57 Beta-ketoacyl synthases active site: G385-V402	CMP-N-acetylneuraminic acid synthetase (g3413320)	HMM BLAST SPScan MOTIFS
2	302	S207 T36 S61 T183 T120 T165 T185 T200 T278 S298 Y148 Y263	N135	Transmembrane domain: L7-W27 Signal peptide: M1-S18 Sialyltransferase family: R74-S128, N203-P248	Alpha 2,6-sialyl-transferase (g1280387)	HMM BLAST SPScan BLOCKS-PFAM
3	578	S28 S49 T77 S114 S228 S259 T291 T388 S404 T412 S452 S133 T243 T412 T433 S507 T564	N90 N112 N289 N450	Signal peptide: M1-G21 Glycosyl hydrolase family: M42-P62, S75-G89, N112-H130, P163-V182, G254-L271, V311-G327, P368-L392, S427-I447	Similar to mannosyl-oligosaccharide alpha-1,2-mannosidase (g3875740)	HMM BLAST SPScan PRINTS

09960729 .032600

09960729 .032806

Table 2 (cont.)

SEQ ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences	Identification	Analytical Methods
4	461	T49 T118 T144 T154 S228 S230 T282 T388 S410 S444 S268 S275 T302 S331 T379 S422 Y339	N38	Glycosyl hydrolases family [BLOCKS]: M1-W24, G46-G79, N85-L119, I361-D372, D399-G426, R439-Y448 Glycosyl hydrolase family 1 [PFAM]: G2-L461 Glycosyl hydrolase family [PRINTS]: T292-I306, I361-P369, D378-F389, K400-W417, R424-A436	Cytosolic beta-glucosidase (g1777770)	BLAST BLOCKS PFAM PRINTS
5	529	T41 T82 T84 T118 S143 T206 T245 S298 T316 T421 S437 S2 T71 S132 S171 T522 Y236	N315 N520	Transmembrane domain: I494-F514 Signal peptide: M1-G21 UDP-glucoronosyl and UDP-glucosyl transferases [MOTIFS]: W356-Q400 UDP-glucoronosyl and UDP-glucosyl transferases [ProfileScan]: N378-T419 UDP-glucoronosyl transferases [BLOCKS]: S34-L56, C127-P167, P190-N213, I255-C282, F295-P344, N350-P394, H449-H488 UDP-glucoronosyl and UDP-glucosyl transferases [PFAM]: G24-K527	UDP-glucuronosyl-transferase (g3135025)	HMM BLAST SPScan MOTIFS ProfileScan BLOCKS PFAM

Table 3

Nucleotide SEQ ID NO:	Nucleotide Range of Useful Fragment	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
6	187-227	Gastrointestinal (0.214) Reproductive (0.205) Nervous (0.162)	Cancer and cell proliferation (0.598) Inflammation (0.225)	PBLUESCRIPT
7	145-195	Reproductive (0.360) Cardiovascular (0.200) Gastrointestinal (0.120) Hematopoietic/Immune (0.120)	Cancer and cell proliferation (0.485) Inflammation (0.273)	PSPORT1
8	487-527	Hematopoietic/Immune (0.200) Reproductive (0.185) Nervous (0.169)	Cancer and cell proliferation (0.500) Inflammation (0.308)	pINCY
9	596-636	Gastrointestinal (0.643) Urologic (0.286) Nervous (0.071)	Cancer and cell proliferation (0.500) Inflammation (0.429)	pINCY
10	1543-1583	Gastrointestinal (0.529) Urologic (0.235) Developmental (0.118)	Cancer and cell proliferation (0.556) Inflammation (0.278)	pINCY

Table 4

Nucleotide SEQ ID NO:	Library	Library Comment
6	U937NOT01	Library was constructed at Stratagene (STR937207), using RNA isolated from the U937 monocyte-like cell line. This line (ATCC CRL1593) was established from malignant cells obtained from the pleural effusion of a 37-year-old Caucasian male with diffuse histiocytic lymphoma.
7	TONGTUT01	Library was constructed using RNA isolated from tongue tumor tissue obtained from a 36-year-old Caucasian male during a hemiglossectomy. Pathology indicated recurrent invasive grade 2 squamous-cell carcinoma.
8	SINTFET03	Library was constructed using RNA isolated from small intestine tissue removed from a Caucasian female fetus, who died at 20 weeks gestation.
9	GBLANOT01	Library was constructed using RNA isolated from diseased gallbladder tissue removed from a 53-year-old Caucasian female during a cholecystectomy. Pathology indicated mild chronic cholecystitis and cholelithiasis with approximately 150 mixed gallstones. Family history included benign hypertension.
10	KIDNTUT15	Library was constructed using RNA isolated from left kidney tumor tissue removed from a 65-year-old Caucasian male during an exploratory laparotomy and nephroureterectomy. Pathology indicated grade 1 renal cell carcinoma within the upper pole of the left kidney. Patient history included malignant melanoma of the abdominal skin, benign neoplasm of colon, cerebrovascular disease, and umbilical hernia. Family history included myocardial infarction, atherosclerotic coronary artery disease, and cerebrovascular disease, and prostate cancer.

Table 5

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Perkin-Elmer Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25: 3389-3402.	*ESTs:* Probability value= 1.0E-8 or less *Full Length sequences*: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183: 63-98; and Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489.	*ESTs:* fasta E value=1.06E-6 *Assembled ESTs*: fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less *Full Length sequences:* fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S and J.G. Henikoff, Nucl. Acid Res., 19:6565-72, 1991. J.G. Henikoff and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37: 417-424.	Score=1000 or greater; Ratio of Score/Strength = 0.75 or larger; and, if applicable, Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol., 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322.	Score=10-50 bits for PFAM hits, depending on individual protein families

09960726 - 050200

Table 5 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25: 217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M. S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M. S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12: 431-439.	Score=3.5 or greater
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch et al. supra; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-5,

b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-5,

c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5, and

d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-5.

2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-5.

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide of claim 3 selected from the group consisting of SEQ ID NO:6-10.

5. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

6. A cell transformed with a recombinant polynucleotide of claim 5.

7. A transgenic organism comprising a recombinant polynucleotide of claim 5.

8. A method for producing a polypeptide of claim 1, the method comprising:

a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

9. An isolated antibody which specifically binds to a polypeptide of claim 1.

10. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10,

b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:6-10,

c) a polynucleotide sequence complementary to a),

d) a polynucleotide sequence complementary to b), and

e) an RNA equivalent of a)-d).

11. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 10.

12. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 10, the method comprising:

a) hybridizing the sample with a probe comprising at least 16 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide, and

b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

13. A method of claim 12, wherein the probe comprises at least 30 contiguous nucleotides.

14. A method of claim 12, wherein the probe comprises at least 60 contiguous nucleotides.

15. A pharmaceutical composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

16. A method for treating a disease or condition associated with decreased expression of functional CME, comprising administering to a patient in need of such treatment the pharmaceutical composition of claim 15.

17. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting agonist activity in the sample.

18. A pharmaceutical composition comprising an agonist compound identified by a method of claim 17 and a pharmaceutically acceptable excipient.

19. A method for treating a disease or condition associated with decreased expression of functional CME, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 18.

20. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

a) exposing a sample comprising a polypeptide of claim 1 to a compound, and

b) detecting antagonist activity in the sample.

21. A pharmaceutical composition comprising an antagonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with overexpression of functional CME, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 21.

23. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 4, the method comprising:

a) exposing a sample comprising the target polynucleotide to a compound, and

b) detecting altered expression of the target polynucleotide.

09980729 .032802

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 9/00

A2

(11) International Publication Number: **WO 00/63351**

(43) International Publication Date: 26 October 2000 (26.10.00)

(21) International Application Number: PCT/US00/10882

(22) International Filing Date: 20 April 2000 (20.04.00)

(30) Priority Data:
60/130,383 21 April 1999 (21.04.99) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 60/130,383 (CIP)
Filed on Not furnished

(71) Applicant *(for all designated States except US)*: INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95054 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monrow Drive #12, Mountain View, CA 94040 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). YANG, Junming [CN/US]; 7125 Bark Lane, San Jose, CA 95129 (US).

(74) Agents: HAMLET–COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: CARBOHYDRATE–MODIFYING ENZYMES

(57) Abstract

The invention provides human carbohydrate–modifying enzymes (CME) and polynucleotides which identify and encode CME. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of CME.

Docket No.: PF-0687 USN

DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe that I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

CARBOHYDRATE-MODIFYING ENZYMES

the specification of which:

/ X / is attached hereto.

/ / was filed on _____ as application Serial No. _____ and if this box contains an X / /, was amended on _____.

/ X / was filed as Patent Cooperation Treaty international application No. PCT/US00/10882 on April 20, 2000, if this box contains an X /_/, was amended on under Patent Cooperation Treaty Article 19 on _____ 2001, and if this box contains an X /_/, was amended on _____.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge my duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim the benefit under Title 35, United States Code, §119 or §365(a)-(b) of any foreign application(s) for patent or inventor's certificate indicated below and of any Patent Cooperation Treaty international applications(s) designating at least one country other than the United States indicated below and have also identified below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application for said subject matter the priority of which is claimed:

Docket No.: PF-0687 USN

Country	Number	Filing Date	Priority Claimed
_____	_____	_____	/_/ Yes /_/ No
_____	_____	_____	/_/ Yes /_/ No

I hereby claim the benefit under Title 35, United States Code, §119(e) of any United States provisional application(s) listed below.

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)
60/130,383	April 21, 1999	Expired

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of Title 35, United States Code §112, I acknowledge my duty to disclose material information as defined in Title 37 Code of Federal Regulations, §1.56(a) which occurred between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

Application Serial No.	Filed	Status (Pending, Abandoned, Patented)

I hereby appoint the following:

Lucy J. Billings	Reg. No. 36,749
Michael C. Cerrone	Reg. No. 39,132
Diana Hamlet-Cox	Reg. No. 33,302
Richard C. Ekstrom	Reg. No. 37,027
Barrie D. Greene	Reg. No. 46,740
Lynn E. Murry	Reg. No. 42,918
Shirley A. Recipon	Reg. No. 47,016
Susan K. Sather	Reg. No. 44,316
Michelle M. Stempien	Reg. No. 41,327
David G. Streeter	Reg. No. 43,168
P. Ben Wang	Reg. No. 41,420

respectively and individually, as my patent attorneys and/or agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith. Please address all communications to:

Docket No.: PF-0687 USN

LEGAL DEPARTMENT
INCYTE GENOMICS, INC.
3160 PORTER DRIVE, PALO ALTO, CA 94304

TEL: 650-855-0555 FAX: 650-849-8886 or 650-845-4166

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

First Joint Inventor:

Full name: Preeti Lal

Signature: Preeti Lal

Date: September 10, 2001

Citizenship: India

Residence: Santa Clara, California CA

P.O. Address: P.O. Box 5142
Santa Clara, California 95056

Second Joint Inventor:

Full name: Henry Yue

Signature: Henry Yue

Date: September 24, 2001

Citizenship: United States

Residence: Sunnyvale, California CA

P.O. Address: 826 Lois Avenue
Sunnyvale, California 94087

Docket No.: PF-0687 USN

Third Joint Inventor:

Full name: Y. Tom Tang

Signature:

Date: Sept 10, 2001

Citizenship: United States

Residence: San Jose, California CA

P.O. Address: 4230 Ranwick Court
San Jose, California 95118

Fourth Joint Inventor:

Full name: ~~Jennifer L. Hillman~~

Signature:

Date: September 21, 2001

Citizenship: United States

Residence: Mountain View, California CA

P.O. Address: 230 Monroe Drive, #17
Mountain View, California 94040

Fifth Joint Inventor:

Full name: Mariah R. Baughn

Signature:

Date: September 5, 2001

Citizenship: United States

Residence: ~~San Leandro, California~~ CA

P.O. Address: 14244 Santiago Road
San Leandro, California 94577

Docket No.: PF-0687 USN

Sixth Joint Inventor:

Full name: Junming Yang

Signature:

Date: September 17, 2001

Citizenship: China

Residence: San Jose, California

P.O. Address: 7125 Bark Lane
San Jose, California 95129

SEQUENCE LISTING

<110> INCYTE PHARMACEUTICALS, INC.
LAL, Preeti
YUE, Henry
TANG, Y. Tom
HILLMAN, Jennifer L.
BAUGHN, Mariah R.
YANG, Junming

<120> CARBOHYDRATE-MODIFYING ENZYMES

<130> PF-0687 PCT

<140> To Be Assigned
<141> Herewith

<150> 60/130,383
<151> 1999-04-21

<160> 10

<170> PERL Program

<210> 1
<211> 434
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 000422CD1

<400> 1

Met	Asp	Ser	Val	Glu	Lys	Gly	Ala	Ala	Thr	Ser	Val	Ser	Asn	Pro	
1				5					10						15
Arg	Gly	Arg	Pro	Ser	Arg	Gly	Arg	Pro	Pro	Lys	Leu	Gln	Arg	Asn	
				20					25						30
Ser	Arg	Gly	Gly	Gln	Gly	Arg	Gly	Val	Glu	Lys	Pro	Pro	His	Leu	
				35					40						45
Ala	Ala	Leu	Ile	Leu	Ala	Arg	Gly	Gly	Ser	Lys	Gly	Ile	Pro	Leu	
				50					55						60
Lys	Asn	Ile	Lys	His	Leu	Ala	Gly	Val	Pro	Leu	Ile	Gly	Trp	Val	
				65					70						75
Leu	Arg	Ala	Ala	Leu	Asp	Ser	Gly	Ala	Phe	Gln	Ser	Val	Trp	Val	
				80					85						90
Ser	Thr	Asp	His	Asp	Glu	Ile	Glu	Asn	Val	Ala	Lys	Gln	Phe	Gly	
				95					100						105
Ala	Gln	Val	His	Arg	Arg	Ser	Ser	Glu	Val	Ser	Lys	Asp	Ser	Ser	
				110					115						120
Thr	Ser	Leu	Asp	Ala	Ile	Ile	Glu	Phe	Leu	Asn	Tyr	His	Asn	Glu	
				125					130						135
Val	Asp	Ile	Val	Gly	Asn	Ile	Gln	Ala	Thr	Ser	Pro	Cys	Leu	His	
				140					145						150
Pro	Thr	Asp	Leu	Gln	Lys	Val	Ala	Glu	Met	Ile	Arg	Glu	Glu	Gly	
				155					160						165
Tyr	Asp	Ser	Val	Phe	Ser	Val	Val	Arg	Arg	His	Gln	Phe	Arg	Trp	
				170					175						180
Ser	Glu	Ile	Gln	Lys	Gly	Val	Arg	Glu	Val	Thr	Glu	Pro	Leu	Asn	
				185					190						195
Leu	Asn	Pro	Ala	Lys	Arg	Pro	Arg	Arg	Gln	Asp	Trp	Asp	Gly	Glu	

```
                200                 205                 210
Leu Tyr Glu Asn Gly Ser Phe Tyr Phe Ala Lys Arg His Leu Ile
                215                 220                 225
Glu Met Gly Tyr Leu Gln Gly Gly Lys Met Ala Tyr Tyr Glu Met
                230                 235                 240
Arg Ala Glu His Ser Val Asp Ile Asp Val Asp Ile Asp Trp Pro
                245                 250                 255
Ile Ala Glu Gln Arg Val Leu Arg Tyr Gly Tyr Phe Gly Lys Glu
                260                 265                 270
Lys Leu Lys Glu Ile Lys Leu Leu Val Cys Asn Ile Asp Gly Cys
                275                 280                 285
Leu Thr Asn Gly His Ile Tyr Val Ser Gly Asp Gln Lys Glu Ile
                290                 295                 300
Ile Ser Tyr Asp Val Lys Asp Ala Ile Gly Ile Ser Leu Leu Lys
                305                 310                 315
Lys Ser Gly Ile Glu Val Arg Leu Ile Ser Glu Arg Ala Cys Ser
                320                 325                 330
Lys Gln Thr Leu Ser Ser Leu Lys Leu Asp Cys Lys Met Glu Val
                335                 340                 345
Ser Val Ser Asp Lys Leu Ala Val Val Asp Glu Trp Arg Lys Glu
                350                 355                 360
Met Gly Leu Cys Trp Lys Glu Val Ala Tyr Leu Gly Asn Glu Val
                365                 370                 375
Ser Asp Glu Glu Cys Leu Lys Arg Val Gly Leu Ser Gly Ala Pro
                380                 385                 390
Ala Asp Ala Cys Ser Thr Ala Gln Lys Ala Val Gly Tyr Ile Cys
                395                 400                 405
Lys Cys Asn Gly Gly Arg Gly Ala Ile Arg Glu Phe Ala Glu His
                410                 415                 420
Ile Cys Leu Leu Met Glu Lys Val Asn Asn Ser Cys Gln Lys
                425                 430
```

<210> 2
<211> 302
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 983984CD1

<400> 2

```
Met Lys Ala Pro Gly Arg Leu Val Leu Ile Ile Leu Cys Ser Val
  1                 5                  10                  15
Val Phe Ser Ala Val Tyr Ile Leu Leu Cys Cys Trp Ala Gly Leu
                 20                  25                  30
Pro Leu Cys Leu Ala Thr Cys Leu Asp His His Phe Pro Thr Gly
                 35                  40                  45
Ser Arg Pro Thr Val Pro Gly Pro Leu His Phe Ser Gly Tyr Ser
                 50                  55                  60
Ser Val Pro Asp Gly Lys Pro Leu Val Arg Glu Pro Cys Arg Ser
                 65                  70                  75
Cys Ala Val Val Ser Ser Ser Gly Gln Met Leu Gly Ser Gly Leu
                 80                  85                  90
Gly Ala Glu Ile Asp Ser Ala Glu Cys Val Phe Arg Met Asn Gln
                 95                 100                 105
Ala Pro Thr Val Gly Phe Glu Ala Asp Val Gly Gln Arg Ser Thr
                110                 115                 120
Leu Arg Val Val Ser His Thr Ser Val Pro Leu Leu Leu Arg Asn
                125                 130                 135
Tyr Ser His Tyr Phe Gln Lys Ala Arg Asp Thr Leu Tyr Met Val
```

```
                  140                 145                 150
Trp Gly Gln Gly Arg His Met Asp Arg Val Leu Gly Gly Arg Thr
                  155                 160                 165
Tyr Arg Thr Leu Leu Gln Leu Thr Arg Met Tyr Pro Gly Leu Gln
                  170                 175                 180
Val Tyr Thr Phe Thr Glu Arg Met Met Ala Tyr Cys Asp Gln Ile
                  185                 190                 195
Phe Gln Asp Glu Thr Gly Lys Asn Arg Arg Gln Ser Gly Ser Phe
                  200                 205                 210
Leu Ser Thr Gly Trp Phe Thr Met Ile Leu Ala Leu Glu Leu Cys
                  215                 220                 225
Glu Glu Ile Val Val Tyr Gly Met Val Ser Asp Ser Tyr Cys Arg
                  230                 235                 240
Glu Lys Ser His Pro Ser Val Pro Tyr His Tyr Phe Glu Lys Gly
                  245                 250                 255
Arg Leu Asp Glu Cys Gln Met Tyr Leu Ala His Glu Gln Ala Pro
                  260                 265                 270
Arg Ser Ala His Arg Phe Ile Thr Glu Lys Ala Val Phe Ser Arg
                  275                 280                 285
Trp Ala Lys Lys Arg Pro Ile Val Phe Ala His Pro Ser Trp Arg
                  290                 295                 300
Thr Glu
```

<210> 3
<211> 578
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2210054CD1

<400> 3

```
Met Pro Phe Arg Leu Leu Ile Pro Leu Gly Leu Leu Cys Ala Leu
  1                 5                  10                  15
Leu Pro Gln His His Gly Ala Pro Gly Pro Asp Gly Ser Ala Pro
                  20                  25                  30
Asp Pro Ala His Tyr Arg Glu Arg Val Lys Ala Met Phe Tyr His
                  35                  40                  45
Ala Tyr Asp Ser Tyr Leu Glu Asn Ala Phe Pro Phe Asp Glu Leu
                  50                  55                  60
Arg Pro Leu Thr Cys Asp Gly His Asp Thr Trp Gly Ser Phe Ser
                  65                  70                  75
Leu Thr Leu Ile Asp Ala Leu Asp Thr Leu Leu Ile Leu Gly Asn
                  80                  85                  90
Val Ser Glu Phe Gln Arg Val Val Glu Val Leu Gln Asp Ser Val
                  95                 100                 105
Asp Phe Asp Ile Asp Val Asn Ala Ser Val Phe Glu Thr Asn Ile
                 110                 115                 120
Arg Val Val Gly Gly Leu Leu Ser Ala His Leu Leu Ser Lys Lys
                 125                 130                 135
Ala Gly Val Glu Val Glu Ala Gly Trp Pro Cys Ser Gly Pro Leu
                 140                 145                 150
Leu Arg Met Ala Glu Glu Ala Ala Arg Lys Leu Leu Pro Ala Phe
                 155                 160                 165
Gln Thr Pro Thr Gly Met Pro Tyr Gly Thr Val Asn Leu Leu His
                 170                 175                 180
Gly Val Asn Pro Gly Glu Thr Pro Val Thr Cys Thr Ala Gly Ile
                 185                 190                 195
Gly Thr Phe Ile Val Glu Phe Ala Thr Leu Ser Ser Leu Thr Gly
                 200                 205                 210
```

```
Asp Pro Val Phe Glu Asp Val Ala Arg Val Ala Leu Met Arg Leu
                215                 220                 225
Trp Glu Ser Arg Ser Asp Ile Gly Leu Val Gly Asn His Ile Asp
                230                 235                 240
Val Leu Thr Gly Lys Trp Val Ala Gln Asp Ala Gly Ile Gly Ala
                245                 250                 255
Gly Val Asp Ser Tyr Phe Glu Tyr Leu Val Lys Gly Ala Ile Leu
                260                 265                 270
Leu Gln Asp Lys Lys Leu Met Ala Met Phe Leu Glu Tyr Asn Lys
                275                 280                 285
Ala Ile Arg Asn Tyr Thr Arg Phe Asp Asp Trp Tyr Leu Trp Val
                290                 295                 300
Gln Met Tyr Lys Gly Thr Val Ser Met Pro Val Phe Gln Ser Leu
                305                 310                 315
Glu Ala Tyr Trp Pro Gly Leu Gln Ser Leu Ile Gly Asp Ile Asp
                320                 325                 330
Asn Ala Met Arg Thr Phe Leu Asn Tyr Tyr Thr Val Trp Lys Gln
                335                 340                 345
Phe Gly Gly Leu Pro Glu Phe Tyr Asn Ile Pro Gln Gly Tyr Thr
                350                 355                 360
Val Glu Lys Arg Glu Gly Tyr Pro Leu Arg Pro Glu Leu Ile Glu
                365                 370                 375
Ser Ala Met Tyr Leu Tyr Arg Ala Thr Gly Asp Pro Thr Leu Leu
                380                 385                 390
Glu Leu Gly Arg Asp Ala Val Glu Ser Ile Glu Lys Ile Ser Lys
                395                 400                 405
Val Glu Cys Gly Phe Ala Thr Ile Lys Asp Leu Arg Asp His Lys
                410                 415                 420
Leu Asp Asn Arg Met Glu Ser Phe Phe Leu Ala Glu Thr Val Lys
                425                 430                 435
Tyr Leu Tyr Leu Leu Phe Asp Pro Thr Asn Phe Ile His Asn Asn
                440                 445                 450
Gly Ser Thr Phe Asp Thr Val Ile Thr Pro Tyr Gly Glu Cys Ile
                455                 460                 465
Leu Gly Ala Gly Gly Tyr Ile Phe Asn Thr Glu Ala His Pro Ile
                470                 475                 480
Asp Pro Ala Ala Leu His Cys Cys Gln Arg Leu Lys Glu Glu Gln
                485                 490                 495
Trp Glu Val Glu Asp Leu Met Arg Glu Phe Tyr Ser Leu Lys Arg
                500                 505                 510
Ser Arg Ser Lys Phe Gln Lys Asn Thr Val Ser Ser Gly Pro Trp
                515                 520                 525
Glu Pro Pro Ala Arg Pro Gly Thr Leu Phe Ser Pro Glu Asn His
                530                 535                 540
Asp Gln Ala Arg Glu Arg Lys Pro Ala Lys Gln Lys Val Pro Leu
                545                 550                 555
Leu Ser Cys Pro Ser Gln Pro Phe Thr Ser Lys Leu Ala Leu Leu
                560                 565                 570
Gly Gln Val Phe Leu Asp Ser Ser
                575
```

<210> 4
<211> 461
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2618358CD1

<400> 4

```
Met Gly Gly Ser Thr Ala Ala Tyr Gln Val Glu Gly Gly Trp Asp
  1               5                  10                  15
Ala Asp Gly Lys Gly Pro Cys Val Trp Asp Thr Phe Thr His Gln
             20                  25                  30
Gly Gly Glu Arg Val Phe Lys Asn Gln Thr Gly Asp Val Ala Cys
         35                  40                  45
Gly Ser Tyr Thr Leu Trp Glu Glu Asp Leu Lys Cys Ile Lys Gln
     50                  55                  60
Leu Gly Leu Thr His Tyr Arg Phe Ser Leu Ser Trp Ser Arg Leu
 65                  70                  75
Leu Pro Asp Gly Thr Thr Gly Phe Ile Asn Gln Lys Gly Ile Asp
             80                  85                  90
Tyr Tyr Asn Lys Ile Ile Asp Asp Leu Leu Lys Asn Gly Val Thr
         95                 100                 105
Pro Ile Val Thr Leu Tyr His Phe Asp Leu Pro Gln Thr Leu Glu
    110                 115                 120
Asp Gln Gly Gly Trp Leu Ser Glu Ala Ile Ile Glu Ser Phe Asp
    125                 130                 135
Lys Tyr Ala Gln Phe Cys Phe Ser Thr Phe Gly Asp Arg Val Lys
    140                 145                 150
Gln Trp Ile Thr Ile Asn Glu Ala Asn Val Leu Ser Val Met Ser
155                 160                 165
Tyr Asp Leu Gly Met Phe Pro Pro Gly Ile Pro His Phe Gly Thr
    170                 175                 180
Gly Gly Tyr Gln Ala Ala His Asn Leu Ile Lys Ala His Ala Arg
    185                 190                 195
Ser Trp His Ser Tyr Asp Ser Leu Phe Arg Lys Lys Gln Lys Gly
    200                 205                 210
Met Val Ser Leu Ser Leu Phe Ala Val Trp Leu Glu Pro Ala Asp
    215                 220                 225
Pro Asn Ser Val Ser Asp Gln Glu Ala Ala Lys Arg Ala Ile Thr
    230                 235                 240
Phe His Leu Asp Leu Phe Ala Lys Pro Ile Phe Ile Asp Gly Asp
    245                 250                 255
Tyr Pro Glu Val Val Lys Ser Gln Ile Ala Ser Met Ser Gln Lys
    260                 265                 270
Gln Gly Tyr Pro Ser Ser Arg Leu Pro Glu Phe Thr Glu Glu Glu
    275                 280                 285
Lys Lys Met Ile Lys Gly Thr Ala Asp Phe Phe Ala Val Gln Tyr
    290                 295                 300
Tyr Thr Thr Arg Leu Ile Lys Tyr Gln Glu Asn Lys Lys Gly Glu
305                 310                 315
Leu Gly Ile Leu Gln Asp Ala Glu Ile Glu Phe Phe Pro Asp Pro
    320                 325                 330
Ser Trp Lys Asn Val Asp Trp Ile Tyr Val Val Pro Trp Gly Val
    335                 340                 345
Cys Lys Leu Leu Lys Tyr Ile Lys Asp Thr Tyr Asn Asn Pro Val
    350                 355                 360
Ile Tyr Ile Thr Glu Asn Gly Phe Pro Gln Ser Asp Pro Ala Pro
    365                 370                 375
Leu Asp Asp Thr Gln Arg Trp Glu Tyr Phe Arg Gln Thr Phe Gln
    380                 385                 390
Glu Leu Phe Lys Ala Ile Gln Leu Asp Lys Val Asn Leu Gln Val
    395                 400                 405
Tyr Cys Ala Trp Ser Leu Leu Asp Asn Phe Glu Trp Asn Gln Gly
    410                 415                 420
Tyr Ser Ser Arg Phe Gly Leu Phe His Val Asp Phe Glu Asp Pro
    425                 430                 435
Ala Arg Pro Arg Val Pro Tyr Thr Ser Ala Lys Glu Tyr Ala Lys
    440                 445                 450
Ile Ile Arg Asn Asn Gly Leu Glu Ala His Leu
    455                 460
```

09980729 .032902

<210> 5
<211> 529
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2912330CD1

<400> 5
Met Ser Met Lys Trp Thr Ser Ala Leu Leu Leu Ile Gln Leu Ser
1 5 10 15
Cys Tyr Phe Ser Ser Gly Ser Cys Gly Lys Val Leu Val Trp Pro
20 25 30
Thr Glu Phe Ser His Trp Met Asn Ile Lys Thr Ile Leu Asp Glu
35 40 45
Leu Val Gln Arg Gly His Glu Val Thr Val Leu Ala Ser Ser Ala
50 55 60
Ser Ile Ser Phe Asp Pro Asn Ser Pro Ser Thr Leu Lys Phe Glu
65 70 75
Val Tyr Pro Val Ser Leu Thr Lys Thr Glu Phe Glu Asp Ile Ile
80 85 90
Lys Gln Leu Val Lys Arg Trp Ala Glu Leu Pro Lys Asp Thr Phe
95 100 105
Trp Ser Tyr Phe Ser Gln Val Gln Glu Ile Met Trp Thr Phe Asn
110 115 120
Asp Ile Leu Arg Lys Phe Cys Lys Asp Ile Val Ser Asn Lys Lys
125 130 135
Leu Met Lys Lys Leu Gln Glu Ser Arg Phe Asp Val Val Leu Ala
140 145 150
Asp Ala Val Phe Pro Phe Gly Glu Leu Leu Ala Glu Leu Leu Lys
155 160 165
Ile Pro Phe Val Tyr Ser Leu Arg Phe Ser Pro Gly Tyr Ala Ile
170 175 180
Glu Lys His Ser Gly Gly Leu Leu Phe Pro Pro Ser Tyr Val Pro
185 190 195
Val Val Met Ser Glu Leu Ser Asp Gln Met Thr Phe Ile Glu Arg
200 205 210
Val Lys Asn Met Ile Tyr Val Leu Tyr Phe Glu Phe Trp Phe Gln
215 220 225
Ile Phe Asp Met Lys Lys Trp Asp Gln Phe Tyr Ser Glu Val Leu
230 235 240
Gly Arg Pro Thr Thr Leu Ser Glu Thr Met Ala Lys Ala Asp Ile
245 250 255
Trp Leu Ile Arg Asn Tyr Trp Asp Phe Gln Phe Pro His Pro Leu
260 265 270
Leu Pro Asn Val Glu Phe Val Gly Gly Leu His Cys Lys Pro Ala
275 280 285
Lys Pro Leu Pro Lys Glu Met Glu Glu Phe Val Gln Ser Ser Gly
290 295 300
Glu Asn Gly Val Val Val Phe Ser Leu Gly Ser Met Val Ser Asn
305 310 315
Thr Ser Glu Glu Arg Ala Asn Val Ile Ala Ser Ala Leu Ala Lys
320 325 330
Ile Pro Gln Lys Val Leu Trp Arg Phe Asp Gly Asn Lys Pro Asp
335 340 345
Thr Leu Gly Leu Asn Thr Arg Leu Tyr Lys Trp Ile Pro Gln Asn
350 355 360
Asp Leu Leu Gly His Pro Lys Thr Lys Ala Phe Ile Thr His Gly
365 370 375
Gly Met Asn Gly Ile Tyr Glu Ala Ile Tyr His Gly Val Pro Met

```
                380                 385                 390
Val Gly Val Pro Ile Phe Gly Asp Gln Leu Asp Asn Ile Ala His
                395                 400                 405
Met Lys Ala Lys Gly Ala Ala Val Glu Ile Asn Phe Lys Thr Met
                410                 415                 420
Thr Ser Glu Asp Leu Leu Arg Ala Leu Arg Thr Val Ile Thr Asp
                425                 430                 435
Ser Ser Tyr Lys Glu Asn Ala Met Arg Leu Ser Arg Ile His His
                440                 445                 450
Asp Gln Pro Val Lys Pro Leu Asp Arg Ala Val Phe Trp Ile Glu
                455                 460                 465
Phe Val Met Arg His Lys Gly Ala Lys His Leu Arg Ser Ala Ala
                470                 475                 480
His Asp Leu Thr Trp Phe Gln His Tyr Ser Ile Asp Val Ile Gly
                485                 490                 495
Phe Leu Leu Thr Cys Val Ala Thr Ala Ile Phe Leu Phe Thr Lys
                500                 505                 510
Cys Phe Leu Phe Ser Cys Gln Lys Phe Asn Lys Thr Arg Lys Ile
                515                 520                 525
Glu Lys Arg Glu
```

<210> 6
<211> 1772
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 000422CB1

<400> 6

```
cggcgggcac tgccaggcgg ggatcgggcg gcgccgagct gaggtggtga gggactagct   60
cccggatgtg gagaagctgg ggagaaggcg tgggaggaag atggactcgg tggagaaggg  120
ggccgccacc tccgtctcca acccgcgggg gcgaccgtcc cggggccggc cgccgaagct  180
gcagcgcaac tctcgcggcg gccagggccg aggtgtggag aagcccccgc acctggcagc  240
cctaattctg gcccggggag gcagcaaagg catccccctg aagaacatta agcacctggc  300
gggggtcccg ctcattggct gggtcctgcg tgcggccctg gattcagggg ccttccagag  360
tgtatgggtt tcgacagacc atgatgaaat tgagaatgtg gccaaacaat ttggtgcaca  420
agttcatcga agaagttctg aagtttcaaa agacagctct acctcactag atgccatcat  480
agaatttctt aattatcata atgaggttga cattgtagga aatattcaag ctacttctcc  540
atgtttacat cctactgatc ttcaaaaagt tgcagaaatg attcgagaag aaggatatga  600
ttctgttttc tctgttgtga gacgccatca gtttcgatgg agtgaaattc agaaaggagt  660
tcgtgaagtg accgaacctc tgaatttaaa tccagctaaa cggcctcgtc gacaagactg  720
ggatggagaa ttatatgaaa atggctcatt ttattttgct aaaagacatt tgatagagat  780
gggttacttg cagggtggaa aaatggcata ctatgaaatg cgagctgaac atagtgtgga  840
tatagatgtg gatattgatt ggcctattgc agagcaaaga gtattaagat atggctattt  900
tggcaaagag aagcttaagg aaataaaact tttggtttgc aatattgatg gatgtctcac  960
caatggccac atttatgtat caggagacca aaaagaaata atatcttatg atgtaaaaga 1020
tgctattggg ataagtttat taaagaaaag tggtattgag gtgaggctaa tctcagaaag 1080
ggcctgttca aagcagacgc tgtcttcttt aaaactggat tgcaaaatgg aagtcagtgt 1140
atcagacaag ctagcagttg tagatgaatg gagaaaagaa atgggcctgt gctggaaaga 1200
agtggcatat cttggaaatg aagtgtctga tgaagagtgc ttgaagagag tgggcctaag 1260
tggcgctcct gctgatgcct gttctactgc ccagaaggct gttggataca tttgcaaatg 1320
taatggtggc cgtggtgcca tccgagaatt tgcagagcac atttgcctac taatggaaaa 1380
ggttaataat tcatgccaaa aatagaaatt agcgtaatat tgagaaaaaa atgatacagc 1440
cttcttcagc cagtttgctt ttatttttga ttaagtaaat tccatgttgt aatgttacag 1500
agagtgtgat ttggtttgtg atatatatat attgtgctct acttttctct ttacgcaaga 1560
taattattta gagactgatt acagtctttc tcagattttt agtaaatgca agtaagaaca 1620
tcatcaaagt tcactttgta ttgtaccctg taaaactgtg tgtttgtgtg ctttcaaaga 1680
tgttgggatt ttatttatct ggggacagtg tgtatggtaa gacatgccct tctattaata 1740
aaactacatt tctcaaactt gaaaaaaaaa aa                               1772
```

<210> 7
<211> 1416
<212> DNA
<213> Homo sapiens

<220>
<221> unsure
<222> 1398
<223> a or g or c or t, unknown, or other

<220>
<221> misc_feature
<223> Incyte ID No: 983984CB1

<400> 7
tctggccgcg cggatcagct tccagcccag tcggcccggc ccgggggcca tggagctccg 60
agcggcggat cgcgagcctc ctgcgaaccc cagcctgcac gcccggttag cattcggccg 120
ggagatgcgg cagtggaatc tggaagggcg gtgaaaaacc tacgtcctgc cctcgcccgg 180
cctctccatt cgtcccccgg gtagagaggt gcccggctcc caccccttcc cagccccagc 240
cctggagaca gcagccccta gactactgag ggacagcgac agcatgaagg ctccgggtcg 300
gctcgtgctc atcatcctgt gctccgtggt cttctctgcc gtctacatcc tcctgtgctg 360
ctgggccggc ctgcccctct gcctggccac ctgcctggac caccacttcc ccacaggctc 420
caggcccact gtgccgggac ccctgcactt cagtggatat agcagtgtgc cagatgggaa 480
gccgctggtc cgcgagccct gccgcagctg tgccgtggtg tccagctccg gccaaatgct 540
gggctcaggc ctgggtgctg agatcgacag tgccgagtgc gtgttccgca tgaaccaggc 600
gcccaccgtg ggctttgagg cggatgtggg ccagcgcagc accctgcgtg tcgtctcaca 660
cacaagcgtg ccgctgctgc tgcgcaacta ttcacactac ttccagaagg cccgagacac 720
gctctacatg gtgtggggcc agggcaggca catggaccgg gtgctcggcg gccgcaccta 780
ccgcacgctg ctgcagctca ccaggatgta ccccggcctg caggtgtaca ccttcacgga 840
gcgcatgatg gcctactgcg accagatctt ccaggacgag acgggcaaga accggaggca 900
gtcgggctcc ttcctcagca ccggctggtt caccatgatc ctcgcgctgg agctgtgtga 960
ggagatcgtg gtctatggga tggtcagcga cagctactgc agggagaaga gccacccctc 1020
agtgccttac cactactttg agaagggccg gctagatgag tgtcagatgt acctggcaca 1080
cgagcaggcg ccccgaagcg cccaccgctt catcactgag aaggcggtct tctcccgctg 1140
ggccaagaag aggcccatcg tgttcgccca tccgtcctgg aggactgagt agcttccgtc 1200
gtcctgccag ccgccatgcc gttgcggagg cctccgggat gtcccatccc aagccatcac 1260
actccacaaa aacatttaat ttatggttcc tgcctcctgc cacgtgctgg gtggacctaa 1320
aggttccttc ccaccccatt ctggccgaca tttggagcca tctcaggcct ccactccctg 1380
agtaattcat ggcatttngg gggctcaccc acctac 1416

<210> 8
<211> 1889
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2210054CB1

<400> 8
gacgtagcgg aagaaaccgc agcagctccc aggatgaact ggttgcagtg gctgctgctg 60
ctgcgggggc gctgagagga cacgagctct atgcctttcc ggctgctcat cccgctcggc 120
ctcctgtgcg cgctgctgcc tcagcaccat ggtgcgccag gtcccgacgg ctccgcgcca 180
gatcccgccc actacaggga gcgagtcaag gccatgttct accacgccta cgacagctac 240
ctggagaatg cctttccctt cgatgagctg cgacctctca cctgtgacgg gcacgacacc 300
tggggcagtt tttctctgac tctaattgat gcactggaca ccttgctgat tttggggaat 360
gtctcagaat tccaaagagt ggttgaagtg ctccaggaca gcgtggactt tgatattgat 420
gtgaacgcct ctgtgtttga aacaaacatt cgagtggtag gaggactcct gtctgctcat 480
ctgctctcca agaaggctgg ggtggaagta gaggctggat ggccctgttc cgggcctctc 540
ctgagaatgg ctgaggaggc ggcccgaaaa ctcctcccag cctttcagac ccccactggc 600

atgccatatg gaacagtgaa cttacttcat ggcgtgaacc caggagagac ccctgtcacc 660
tgtacggcag ggattgggac cttcattgtt gaatttgcca ccctgagcag cctcactggt 720
gacccggtgt tcgaagatgt ggccagagtg gctttgatgc gcctctggga gagccggtca 780
gatatcgggc tggtcggcaa ccacattgat gtgctcactg gcaagtgggt ggcccaggac 840
gcaggcatcg gggctggcgt ggactcctac tttgagtact tggtgaaagg agccatcctg 900
cttcaggata agaagctcat ggccatgttc ctagagtata acaaagccat ccggaactac 960
acccgcttcg atgactggta cctgtgggtt cagatgtaca aggggactgt gtccatgcca 1020
gtcttccagt ccttggaggc ctactggcct ggtcttcaga gcctcattgg agacattgac 1080
aatgccatga ggaccttcct caactactac actgtatgga agcagtttgg ggggctcccg 1140
gaattctaca acattcctca gggatacaca gtggagaagc gagagggcta cccacttcgg 1200
ccagaactta ttgaaagcgc aatgtacctc taccgtgcca cgggggatcc caccctccta 1260
gaactcggaa gagatgctgt ggaatccatt gaaaaaatca gcaaggtgga gtgcggattt 1320
gcaacaatca aagatctgcg agaccacaag ctggacaacc gcatggagtc gttcttcctg 1380
gccgagactg tgaaatacct ctacctcctg tttgacccaa ccaacttcat ccacaacaat 1440
gggtccacct tcgacacggt gatcaccccc tatggggagt gcatcctggg ggctgggggg 1500
tacatcttca acacagaagc tcaccccatc gaccctgccg ccctgcactg ctgccagagg 1560
ctgaaggaag agcagtggga ggtggaggac ttgatgaggg aattctactc tctcaaacgg 1620
agcaggtcga aatttcagaa aaacactgtt agttcggggc catgggaacc tccagcaagg 1680
ccaggaacac tcttctcacc agaaaaccat gaccaggcaa gggagaggaa gcctgccaaa 1740
cagaaggtcc cacttctcag ctgccccagt cagcccttca cctccaagtt ggcattactg 1800
ggacaggttt tcctagactc ctcataacca ctggataatt tttttatttt tatttttttg 1860
aggctaaact ataataaatt gcttttggt 1889

<210> 9
<211> 2135
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2618358CB1

<400> 9
ctcctctggc caaagggtgc tctgcttctg gcagctgaag atcccagtag acagcttctt 60
aaaccatggc tttccctgca ggatttggat gggcggcagc actgcagctt atcaagtaga 120
aggaggctgg gatgcagatg gaaaaggccc ttgtgtctgg gacacattta ctcatcaggg 180
aggagagaga gttttcaaga accagactgg cgatgtagct tgtggcagct acactctgtg 240
ggaggaagat ttgaaatgta tcaaacagct tggattgact cattaccgct tctctctttc 300
ctggtcacgt ctgttacctg atgggacgac aggtttcatc aaccagaaag gaattgatta 360
ttacaacaag atcatcgatg atttgttaaa aaatggggtt actcccattg tgaccctcta 420
ccactttgat ttgcctcaga ctttagaaga ccaaggaggt tggttgtcag aggcaatcat 480
tgaatccttt gacaaatatg ctcagttttg cttcagtacc tttggggatc gtgtcaagca 540
gtggatcacc ataaatgaag ctaatgttct ttctgtgatg tcatatgact taggtatgtt 600
tcctccgggt atccctcact ttgggactgg aggttatcag gcagctcata atttgattaa 660
ggctcatgcc agatcctggc acagctatga ttccttattt cgaaaaaagc agaaaggtat 720
ggtgtctcta tcacttttg cggtctggtt ggaaccagca gatcccaact cagtgtctga 780
ccaggaagct gctaaaagag ccatcacttt ccatctggat ttatttgcta aacccatatt 840
catcgatggt gattatcctg aagttgtcaa gtctcagatt gcctccatga gtcaaaagca 900
aggctatcca tcatcgaggc ttccagaatt cactgaagaa gagaagaaaa tgatcaaagg 960
cactgctgat ttttttgctg tgcaatatta tacaactcgc ttaatcaagt accaggagaa 1020
caagaaagga gaactaggta ttctccagga tgcggaaatt gaatttttc cagatccatc 1080
ttggaaaaat gtggattgga tctacgtggt accatgggga gtatgtaaac tactgaaata 1140
tattaaggat acatataata accctgtaat ttacatcact gagaatgggt ttccccagag 1200
tgacccagcg cctcttgatg acactcaacg ctgggagtat ttcagacaaa catttcagga 1260
actgttcaaa gctatccaac ttgataaagt caatcttcaa gtatattgtg catggtctct 1320
tctggataac tttgagtgga accagggata cagcagccgg tttggtctct tccacgttga 1380
ttttgaagac ccagctagac cccgagtccc ttacacatcg gccaaggaat atgccaagat 1440
catccgaaac aatggccttg aagcacatct gtaggcaaga tggctgagaa atacaggaga 1500
ggcgtctgct tttggaaagg aaatctgctt tggtgatgat ctttcaggca atctcaactt 1560
acttctttaa tcaacattta atatcaatgg atctgtgatt aaaaggtctg aatatgtaat 1620
gcctcgtgaa gtatttaata atggccttta tttgtatttg gatcaatgag gtttttaaaa 1680

aaaatggaag agaaaaccac taaccttgat ttttgtattg caaaatcaga tagacctgga 1740
aacataaatt taaatcctta gacatttttc tagaaaaaaa tgcaaagttt ataaagatga 1800
tacaaccatg atttgcaact gtaacaggag accatttatt ataagcgtac ctgtttgtga 1860
acttaattat tctgattcca taagctgttt ttgcttaggt gatccactgc catgtgatcc 1920
ataatttttc tacataaaaa atcaaagtta aaagtcacat tatacagtta tgcattcatt 1980
tcaacaaaat agtgaattga taatctactt gttaatatat tcggcccata ttttgtgtgt 2040
ttggacaagt acatctccct tttgcctaat gaacttttga aaaataataa aataatagaa 2100
taaattagac tttgaatggc aaaaaaaaaa aaaaa 2135

<210> 10
<211> 1650
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 2912330CB1

<400> 10
agcaactgga aaacaagcat tgcattgcat caggatgtct atgaaatgga cttcagctct 60
tctcctgata cagctgagct gttactttag ctctgggagt tgtggaaagg tgctggtgtg 120
gcccacagaa ttcagccact ggatgaatat aaagacaatc ctggatgaac ttgtccagag 180
aggtcatgag gtgactgtat tggcatcttc agcttccatt tctttcgatc ccaacagccc 240
atctactctt aaatttgaag tttatcctgt atctttaact aaaactgagt ttgaggatat 300
tatcaagcag ctggttaaga gatgggcaga acttccaaaa gacacatttt ggtcatattt 360
ttcacaagta caagaaatca tgtggacatt taatgacata cttagaaagt tctgtaagga 420
tatagtttca aataagaaac ttatgaagaa actacaggag tcaagatttg atgttgttct 480
tgcagatgct gttttccccc ttggtgagct gctggccgag ttacttaaaa tacccctttgt 540
ctacagcctc cgcttctctc ctggctacgc aattgaaaag catagtggag gacttctgtt 600
ccctccttcc tatgtgcctg ttgttatgtc agaactaagt gaccaaatga ctttcataga 660
gagggtaaaa aatatgatct atgtgcttta ttttgaattt tggttccaaa tatttgacat 720
gaagaagtgg gatcagttct acagtgaagt tctaggaaga cccactacgt tatctgagac 780
aatggcaaaa gctgacatat ggcttattcg aaactactgg gattttcaat ttcctcaccc 840
actcttacca aatgttgagt tcgttggagg actccactgc aaacctgcca aacccctacc 900
gaaggaaatg gaagagtttg tccagagctc tggagaaaat ggtgttgtgg tgttttctct 960
ggggtcgatg gtcagtaaca cgtcagaaga aagggccaat gtaattgcat cagcccttgc 1020
caagatccca caaaaggttc tgtggagatt tgatgggaat aaaccagata ctttaggact 1080
caatactcgg ctgtacaagt ggatacccca gaatgatctt cttggtcatc ccaaaaccaa 1140
agcttttatc actcatggtg gaatgaatgg gatctatgaa gctatttacc atggggtccc 1200
tatggtggga gttcccatat ttggtgatca gcttgataac atagctcaca tgaaggccaa 1260
aggagcagct gtagaaataa acttcaaaac tatgacaagc gaagatttac tgagggcttt 1320
gagaacagtc attaccgatt cctcttataa agagaatgct atgagattat caagaattca 1380
ccatgatcaa cctgtaaagc ccctagatcg agcagtcttc tggatcgagt ttgtcatgcg 1440
ccacaaagga gccaagcacc tgcgatcagc tgcccatgac ctcacctggt tccagcacta 1500
ctctatagat gtgattgggt tcctgctgac ctgtgtggca actgctatat tcttgttcac 1560
aaaatgtttt ttattttcct gtcaaaaatt taataaaact agaaagatag aaaagaggga 1620
atagatcttt ccaaattcaa gaaagacctg 1650

09980729.032802

United States Patent & Trademark Office

Office of Initial Patent Examination -- Scanning Division

Application deficiencies found during scanning:

☑ Page(s) _____ of Preliminary amendment ~~were~~ was not present for scanning. (Document title)

☐ Page(s) _____ of _____ were not present for scanning. (Document title)

☐ ***Scanned copy is best available.***

[redacted]